THE

AMERICAN NATURALIST

VOL. XL | November, 1906 | No. 479

VARIATION IN THE NUMBER OF SEEDS OF THE LOTUS

RAYMOND PEARL

IN HIS *Mutationstheorie* (vol. 1, p. 112) de Vries puts at the head of a list of topics for further investigation by the student of variation the following sentence: "Das Quetelet'sche Gesetz bedarf immer weiterer Beispiele; die Zahl dieser kann nie gross genug werden." In view of this statement from so distinguished an investigator of the problems of evolution, I venture to publish some material on variation in Nelumbium which has been in my notes for some years and which has frequently been used as an illustration in classroom lectures in biometry. As will be seen in what follows, this material conforms very closely to the normal or Gaussian law in the distribution of its variates; much more closely in point of fact than do many cases which have commonly been cited as typical illustrations of that law.

In marshy situations at many points about the shores of the western part of Lake Erie the common lotus, *Nelumbium luteum* Willd., grows in great abundance.[1] Especially in a strip of water known locally as "Black Channel," which connects Sandusky Bay with the lake, does this plant flourish. Many acres of water are literally covered with its leaves. Pieters (*loc. cit.*, p. 66) says of the growth of Nelumbium in this region: "The immense yellow flowers rising just above the great dark-green standing leaves and the water covered with huge floating pads make this the most striking formation of the swamp. The Nelumbium grows in from

[1] *Cf.* Pieters, A. J. "The Plants of Western Lake Erie, with Observations on their Distribution." *Bull. U. S. Fish Comm. 1901*, pp. 57–79.

2 to 4 feet of water or stray plants may be found in less than 2 feet. Many of the floating leaves were 20 to 24 inches across and the standing ones not much smaller. At Upper Sandusky Bay I found a floating leaf 26 inches in diameter and another with a petiole more than 5 feet in length. Both at Sandusky Bay and along the Portage River the acreage of Nelumbium was greater than at East Harbor, but nowhere did the plants present a more vigorous growth or so magnificent an appearance."

The large ovoid seeds of this plant are borne in pockets scattered

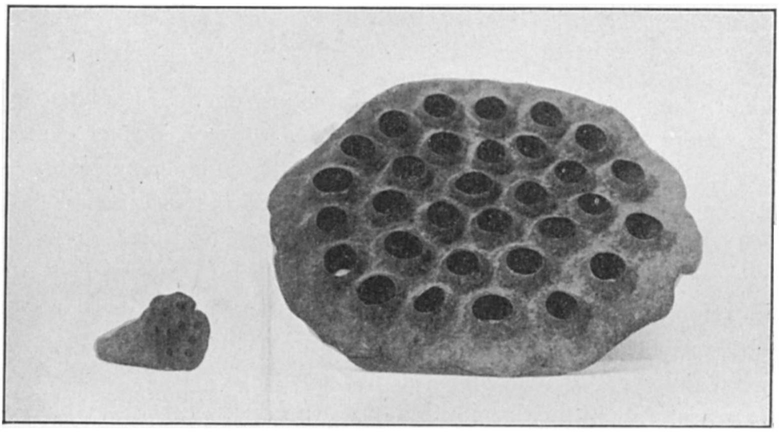

FIG. 1.— Showing the general form of the capsule and arrangement of the seeds in Nelumbium. The two capsules shown in this photograph represent the extremes of variation in the sample; the capsule on the left bore 9 seeds, and the one on the right 39. In the photograph both are reduced below actual size to the same degree.

over the flat, upper surface of the conical seed capsule. After the flower has been shed the ends of the seeds are seen projecting from these pockets. The form of the capsule and the arrangement of the seeds are shown in the accompanying photographs (Figs. 1, 2, and 3), for the preparation of which I am indebted to Miss Frances J. Dunbar.

It is the purpose of the present paper to set forth the results of a study of the variation in the number of seeds to the flower (or the capsule) in this plant. At the end of the flowering season in the summer of 1902 a series of 1410 seed capsules was collected at random from the Black Channel fields in Sandusky Bay. A count was made of the number of seeds in each of these capsules and the records so obtained form the basis of this paper.

The raw data are exhibited in Table 1.

TABLE 1

Frequency Distribution of Number of Seeds in Nelumbium

Number of Seeds per Capsule	Frequency	Number of Seeds per Capsule	Frequency	Number of Seeds per Capsule	Frequency
9	1	20	60	31	45
10	0	21	101	32	34
11	0	22	111	33	21
12	2	23	113	34	13
13	2	24	114	35	11
14	1	25	107	36	7
15	13	26	137	37	2
16	11	27	120	38	1
17	30	28	101	39	1
18	41	29	90		
19	58	30	62	Total	1410

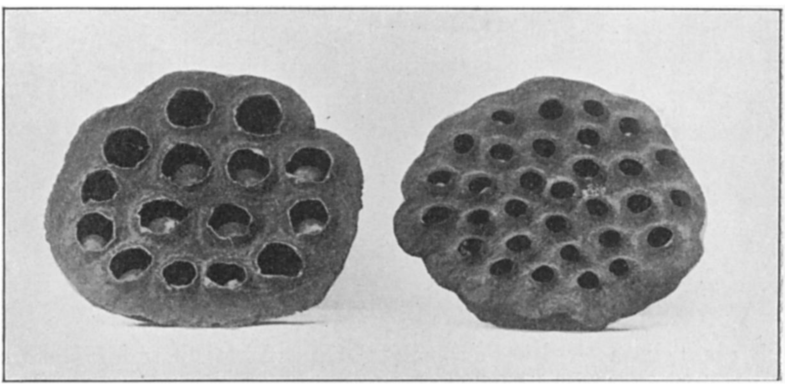

FIG. 2.— Showing two capsules of almost exactly the same size but bearing widely different numbers of seeds, the one on the left having 15 seeds while the other has 31. It should be noted that the openings of the seed pockets have been enlarged with a knife in the specimen on the left. The normal aspect of the capsule top is shown in the right-hand specimen.

The chief physical constants for this distribution are given in Table 2.

TABLE 2

Constants for Variation in Seed Number in Nelumbium

Mean	24.874 ± .078
Standard Deviation	4.339 ± .055
Coëfficient of Variation	17.445 ± .162

It will be noted that the distribution as a whole is quite symmetrical. The relative variability, as measured by the coëfficient of variation, is of the same general order of magnitude as has been found

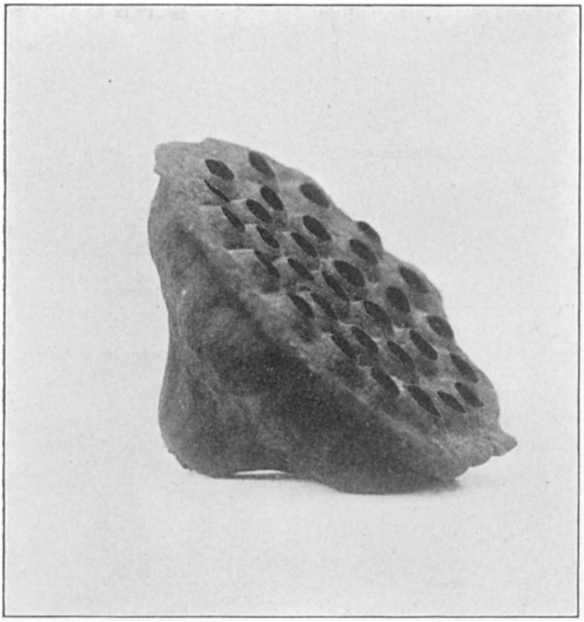

FIG. 3.— A large, fully developed capsule seen from the side.

in plant characters by other workers. In order to determine whether or not the variation of the character under consideration follows the normal law within the limits of the errors of random sampling we must examine the values of the analytical constants, which define the character of a frequency distribution, in comparison with their probable errors. Using Sheppard's corrections for the moments I find the values given in Table 3. The unit for the moments is 1 seed.

TABLE 3

Analytical Constants for Variation in Nelumbium

Constant	Value
μ_2	18.8307
μ_3	2.4675
μ_4	1022.5949
β_1	0.0009
$\sqrt{\beta_1}$	0.0302
β_2	2.8838
β_2-3	—0.1162
κ_1	—0.2351
κ_2	—0.0029
Skewness	0.0164
Modal Divergence	0.0712
Standard Deviation	4.3394
Mean	24.8745
Mode	24.8033

Further, we have the following values for the probable errors of the chief constants concerned in testing whether the distribution sensibly deviates from the normal law. It will be understood that these are the values of the probable errors for the normal curve.

Probable error of skewness $= \pm\ 0.0220$
" " " $\sqrt{\beta_1} = \pm\ 0.0440$
" " " $\beta_2 = \pm\ 0.0880$
" " " modal divergence $= \pm\ 0.0955$

We see at once that neither the skewness, the difference between the mean and the mode, nor $\sqrt{\beta_1}$, are sensibly different from what they would be for an absolutely normal distribution. In the case of each of these constants the theoretical value for a normal curve is zero. The values found from the actual statistics in this reasonably large sample differ from zero by less than the probable errors. Hence we may conclude that in respect to number of seeds per capsule Nelumbium varies symmetrically about the mean (which of course coincides with the modal) condition. A half of the capsules bear less than the typical number of seeds, and a half more than the typical number. Turning to the quantity β_2-3, which measures the degree of flatness at the top of the curve, or, as it has been called by Pearson,[1] the *kurtosis*, the case is somewhat different. Theoretically the normal curve is mesokurtic, or $\beta_2 - 3 = 0$. Now in the present case $\beta_2 - 3$ differs from zero by more than its probable error. The deviation is less than twice the probable error of β_2, so cannot be considered as significant on this basis. As we shall see, however, we get a somewhat better fit to the data given by the *actual sample* if we use a curve which takes into account this deviation from the mesokurtic condition of the normal curve. In so far, however, as we may infer from the sample regarding the conditions in the general population from which the sample is taken, we can conclude with a high degree of probability that *in the variation in number of seeds per capsule Nelumbium follows the normal law of errors.*

From the values of κ_1 and κ_2 given in Table 3 we see that whatever deviation from normality exists, is in the direction of a curve of Type 1. In order to compare the graduation given by a normal and a skew curve, I have fitted both types of curve to the data. The equation to the normal curve is

$$y = 129.6271 \; e^{-\frac{x^2}{37.6614}}$$

while the equation to the Type 1 curve is

$$y = 127.6421 \left(1 + \frac{x}{28.7285}\right)^{21.7365} \left(1 - \frac{x}{32.1508}\right)^{24.3259}.$$

Calculating out the ordinates of these two curves corresponding to the different numbers of seeds, we have the results shown in Table 4.

TABLE 4

Comparison of Observations and Fitted Curves

Number of Seeds per Capsule	Observed Frequency	Ordinates of Normal Curve	Ordinates of Type 1 Curve
9	1	.2	.06
10	0	4	.2
11	0	8	.5
12	2	1.6	1.2
13	2	3.1	2.6
14	1	5.6	5.2

[1] *Biometrika*, vol. 4, p. 173.

TABLE 4 (*continued*)

Number of Seeds per Capsule	Observed Frequency	Ordinates of Normal Curve	Ordinates of Type 1 Curve
15	13	9.7	9.5
16	11	16.0	16.2
17	30	25.0	25.7
18	41	37.0	38.3
19	58	51.8	53.6
20	60	69.0	70.8
21	101	87.0	88.4
22	111	104.1	104.7
23	113	118.1	117.7
24	114	127.0	125.6
25	107	129.6	127.4
26	137	125.3	123.3
27	120	115.0	113.3
28	101	100.0	99.0
29	90	82.5	82.3
30	62	64.5	65.0
31	45	47.9	48.7
32	34	33.7	34.6
33	21	22.5	23.2
34	13	14.2	14.7
35	11	8.5	8.8
36	7	4.9	4.9
37	2	2.6	2.6
38	1	1.3	1.2
39	1	.7	.6

Of course to get absolute accuracy, areas instead of ordinates should be compared with the observed frequencies, but inasmuch as the number of groups is here large, the error made by comparing ordinates will not be serious.

The frequency polygon and fitted curves are shown in Fig. 4.

The fit is seen to be excellent in the case of both the curves, but the slight superiority of the Type 1 curve is apparent. The difference, as has been pointed out above (p. 762), between this and the normal curve is not significant. The greatest discrepancy between the observations and the curves is in the region about the mode. I am unable to account for the curious irregularity in the

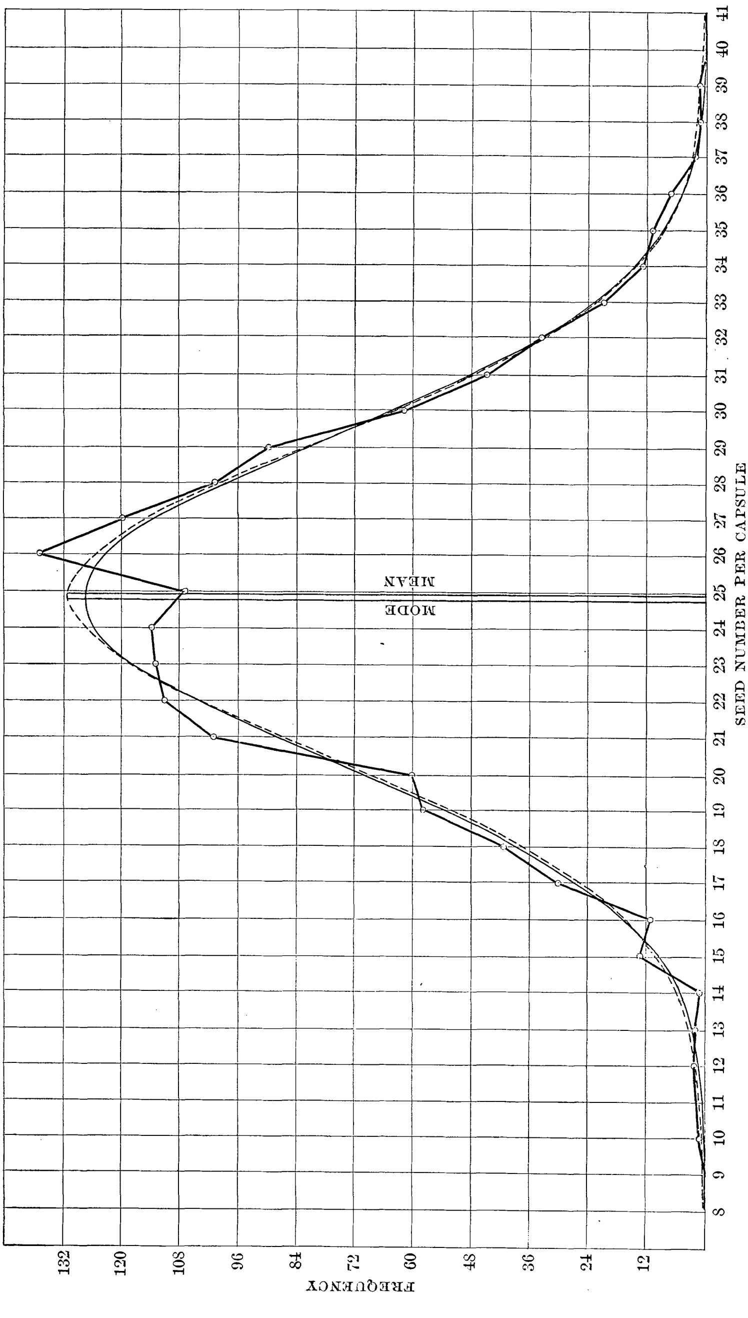

Fig. 4.— Diagram showing variation in the lotus. The abscissæ give seed number and the ordinates frequencies, o——o observations; - - - -, normal curve; ——, type 1 curve,

observation polygon in this region except as a result of random sampling.

The fact that this distribution approaches very closely to the normal type is indicated by the value obtained for the theoretical range of variation when a Type 1 curve is used. It will be recalled that this type of curve has the range limited in both directions, while the normal curve has an infinite range. Using the values of the moments given in Table 3, I find for the Type 1 curve:—

Total range = 60.8794
Lower limit of range = −3.9252
Upper " " " = 56.9541

It is clear that the theoretical range greatly overestimates the observed. Of course the start at −4 seeds appears at first sight to be an absurdity, but it must be remembered that this value is subject to a considerable probable error, and that it is possible to get as great an extension as this of the range of the theoretical curve below zero as a result merely of random sampling. Furthermore it must be admitted that while the upper limit of the range at 57 seeds seems very improbable, yet, for anything we know to the contrary, it is not impossible.[1] In general it is clear from this case that as the Type 1 curve approaches the normal its range becomes greatly extended.

There is one further point regarding this material to which attention should be called, namely, the bearing of the results on the question of the distribution of fecundity. It is evident that the number of seeds borne by a plant is the measure of its fecundity. In considering data like those here presented the question at once arises as to whether each different class of capsules contributes its proportionate share in the total number of seeds available for the propagation of a succeeding generation. A moment's consideration shows that this cannot be the case in Nelumbium. The figures given in Table 5 demonstrate this. To avoid the possibility of misunderstanding, the manner in which this table is formed may

[1] Since writing the above I have seen some actual statistics of variation in seed number in the lotus in which the upper limit of the observed range is 42 seeds, showing a tendency in the direction predicted by the theoretical curve.

be stated briefly. The figures in the second column were obtained by multiplying the number of seeds in a given capsule by the frequency with which that class of capsule occurred in the sample. The third column gives the same data reduced to *per mille* proportions.

TABLE 5

Total Number of Seeds borne by Capsules of Different Sizes

Capsule Class (Seeds per Capsule)	Total Number of Seeds borne in all Capsules of Designated Class	Per Mille Number of Seeds borne in all Capsules of Designated Class
9	9	0.26
10	0	0
11	0	0
12	24	0.68
13	26	0.74
14	14	0.40
15	195	5.56
16	176	5.02
17	510	14.54
18	738	21.04
19	1102	31.42
20	1200	34.21
21	2121	60.47
22	2442	69.63
23	2599	74.10
24	2736	78.02
25	2675	76.27
26	3562	101.56
27	3240	92.38
28	2828	80.63
29	2610	74.42
30	1860	53.03
31	1395	39.77
32	1088	31.02
33	693	19.76
34	442	12.60
35	385	10.98
36	252	7.19
37	74	2.11
38	38	1.08
39	39	1.11
Total	35,073	1000.00

From this table we see that, in round numbers, 1400 capsules produce 35,000 seeds. Further, it is clear that the different classes of capsules do not contribute in proportion to their frequency of occurrence to the total seed number. Thus, for example, a reference to Table 1 shows that capsules with 21 seeds each and capsules with 28 seeds each occur with equal frequency in our sample. But obviously the latter will contribute more to the total number of seeds. As a matter of fact the 28-seed capsules contribute 81 per thousand of the total number of seeds, as against 60 per thousand of the 21-seed capsules. Taking the data as a whole I find by a very simple calculation that:

(*a*) Capsules with *fewer* than the median number of seeds bear altogether 15066.325 seeds, or 42.96 percent of the total number.

(*b*) Capsules with *more* than the median number of seeds bear altogether 20006.675 seeds, or 57.04 percent of the total number. In other words 50 percent of the capsules produce 57 percent of the seeds, or, put in still another way, one half of the heads bears 14 percent more of the total number of seeds than does the other half. This result is, of course, an obviously necessary arithmetical consequence of the symmetry of the capsule distribution, yet it is a point which is frequently overlooked. A symmetrical distribution of the individuals of a population with respect to some measure of fecundity does not mean that the contributions of these individuals to the next generation even before selection will be represented by a symmetrical distribution. The very fact that the original distribution is symmetrical necessitates the contrary relation.

The results with reference to the proportionate contributions of the different classes of heads to the total seed number show the conditions before elimination begins. Many of the 35,000 seeds were undoubtedly incapable of germination, and after germination many more would be eliminated before reaching maturity. As to the distribution of the eliminating factors acting in the case of the lotus we know nothing. What I wish to emphasize here is that out of the total number of seeds before elimination begins, 57 percent are the product of one half of the parent heads and only 43 percent the product of the other half.

The results of this study may be summarized briefly as follows:

(1) In the variation in respect to number of seeds per capsule *Nelumbium luteum* follows very closely the normal or Gaussian law of the distribution of errors.

(2) Place constants are given for the designated character in the form unit of Nelumbium growing in Sandusky Bay.

(3) From the fact that the frequency distribution of the capsules in respect to seed number is symmetrical about the mean it follows that one half of the whole number of capsules bears 14 percent more of the total number of seeds available for a new generation than does the other half of the capsules.

ZOÖLOGICAL LABORATORY
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PA.